Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 68 (2007) 2163–2175
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

On the processing of metabolic information through metabolite–gene communication networks: An approach for modelling causality

Jedrzej Szymanski [a], Monika Bielecka [a], Fernando Carrari [a,b], Alisdair R. Fernie [a], Rainer Hoefgen [a], Victoria J. Nikiforova [a,c,*]

[a] *Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*
[b] *Instituto de Biotechnología, CICVyA, Instituto Nacional de Technología Agrícola Argentina, Buenos Aires, Argentina*
[c] *Timiryazev Institute of Plant Physiology, Russian Academy of Sciences, Moscow 127276, Russia*

Received 19 January 2007; received in revised form 5 April 2007
Available online 4 June 2007

Abstract

Gene-metabolite correlation networks of three independent biological systems were interrogated using an approach to define, and subsequently model, causality. The major goal of this work was to analyse how information from those metabolites, that displayed a rapid response to perturbation of the biological system, is processed through the response network to provide signal-specific adaptation of metabolism. For this purpose, comparison of network topologies was carried out on three different groups of system elements: transcription factors, other genes and metabolites, with special emphasis placed on those features which are possible sites of metabolic regulation or response propagation. The degree of connectivity in all three analysed gene-metabolite networks followed power-law and exponential functions, whilst a comparison of connectivities of the various cellular entities suggested, that metabolites are less involved in the regulation of the sulfur stress response than in the ripening of tomatoes (in which metabolites seem to have an even greater regulatory role than transcription factors). These findings reflect different degree of metabolic regulation for distinct biological processes. Implementing causality into the network allowed classification of metabolite-gene associations into those with causal directionality from gene to metabolite and from metabolite to gene. Several metabolites were positioned relatively early in the causal hierarchy and possessed many connections to the downstream elements. Such metabolites were considered to have higher regulatory potential. For the biological example of hypo-sulfur stress response in Arabidopsis, the highest regulatory potential scores were established for fructose and sucrose, isoleucine, methionine and sinapic acid. Further developments in profiling techniques will allow greater cross-systems comparisons, necessary for reliability and universality checks of inferred regulatory capacities of the particular metabolites.

Keywords: *Solanum lycopersicum*; Solanaceae; Tomato; *Arabidopsis thaliana*; Brassicaceae; Sulfur starvation; Fruit ripening; Correlation network; Causality; Metabolic regulation

1. Introduction

A living organism consists of many highly diverse molecular entities (such as genes, proteins, metabolites) organized in a functional dynamic system. This system must be capable of simultaneously maintaining homeostasis and reacting to changes within its environment. In order to perform both functions, the systemic response to perturbation constitutes a branched chain of consecutive changes of cellular entities. Information concerning these changes propagates along this chain, forming a dense network of interactions. Given that changes in states of downstream elements occur as a result of preceding upstream changes, it follows that infor-

* Corresponding author. Address: Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany. Tel.: +49 331 5678254; fax: +49 331 5678250.
E-mail address: nikiforova@mpimp-golm.mpg.de (V.J. Nikiforova).

0031-9422/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2007.04.017

mation processing induced by a perturbation (signal) is causally directed – starting from an exciter (a cause) and its perception through transduction and transformation to an endpoint response (an effect). For biologists who attempt to decipher these chains of response reactions and to elucidate regulatory points within them, the intrinsic causality of information processing networks represents a useful framework that, in combination with empirical biological knowledge, can be used in the analysis of biological reasoning. In order to attempt such evaluation two basic features are required within the dataset under consideration: (i) the connections between recorded changes reflecting their putative mutual dependencies must be determinable and (ii) some biological knowledge concerning temporal points of excitation (or final effect) must be available. For this reason kinetic experiments appear to provide the richest source of information for such studies. Time series of transcript and metabolite profiling data, allowing cross-correlation analysis and obtained in the experiments with plants, when the response was induced by a treating agent recognized as a response exciter, clearly satisfy both of the above mentioned prerequisites. The present study is focused on a particular aspect of the information processing in plants, which is the transmission of information from a metabolite to a gene. To access this question, we decipher causality in the integrated gene-metabolite correlation networks.

Whilst the **integrated analysis of metabolite and gene expression profiles** is recognized nowadays as a powerful tool for biotechnology (Fernie et al., 2005), and is increasingly useful in gene annotation studies (Tohge et al., 2005), relatively little attempt has been made to decipher information flow across and between these molecular entities. Initial experiments in which transcript and metabolic profiles were analysed in parallel revealed important co-regulatory behaviour between different molecular species (Askenazi et al., 2003; Urbanczyk-Wochniak et al., 2003). Subsequently from more detailed analysis of the systemic response of *Arabidopsis thaliana* to sulfur deficiency it was possible to identify clusters of co-regulated genes and metabolites, that were interpretable on the basis of a priori biological knowledge (Hirai et al., 2004, 2005). Similarly, several aspects of the regulation of metabolism were revealed following the interrogation of the integrated gene-metabolite network of tomato fruit ripening (Carrari et al., 2006) and folate responsiveness in mice (Ernest et al., 2006). When taken together, data from these examples (and several subsequent studies: Scheible et al., 2004; Osuna et al., 2007) suggest that there is often a considerable delay in the temporal response of metabolites with respect to transcripts. These time response lags have to be taken into account when biological reasoning in the informational exchange between genes and metabolites is accessed.

Approaches to infer **causality** have been developed and applied for gene co-expression networks. These are generally based on dynamic measurements of response which yield hierarchical information about causal relationships within networks. Such analysis has been performed to dissect gene networks following hormone and insulin signaling data (Kam, 2002), whilst information regarding the time lag between species at which the highest correlation was found has also been used as a method to infer causality (D'haeseleer et al., 2000). However, such reasoning cannot be directly extrapolated from pure gene co-expression networks to gene–metabolite correlation networks due to the multi-scale nature of metabolic and gene expression responses. To address to this problem, we previously elaborated a new approach for implementing causal directionality into gene-metabolite correlation networks with the use of the a priori knowledge on the molecule, which excites the systems response and can thus be considered as a 'cause' (Nikiforova et al., 2005b). In such networks, propagation of the information flow from the exciter to physiological endpoints through alterations in gene expressions and metabolite concentrations can be followed. Keeping the strict statistical regime about significance of observed patterns in correlation networks, from this analysis we can conclude biologically meaningful information, even if observed interactions are putative or indirect.

In the current work we apply this approach to extract from gene–metabolite correlation network causally directed associated pairs metabolite-to-gene and estimate the putative regulatory potential of the metabolites by means of comparative network topology analysis. Three exemplary gene–metabolite networks (sulfur stress response in *A. thaliana*, application of inhibitors of folate biosynthesis to *A. thaliana*, and the tomato fruit ripening process) are considered. In an attempt to identify putative regulators of the adaptive response, capable to influence inter alia gene expression, we distinguish those pairs of gene–metabolite associations, in which dependency is putatively directed from metabolite to gene. Three groups of biosystem elements with regulatory potential were considered: metabolites, transcription factors as known regulators of gene expression and other genes. Putative expressional regulation activity of metabolites, which we put in the focus of the current research, has been recently recognized (Tucker and Breaker, 2005; Grundy and Henkin, 2006; Ladurner, 2006) and is not yet well understood.

Data from the two successful implementations are discussed in the context of current models of metabolic regulation of gene expression. In addition, the reasons behind the inability to implement causality in the third dataset, as well as current limitations and future prospects of this approach, are discussed.

2. Results and discussion

2.1. Suitable datasets for reconstruction of causally directed gene–metabolite correlation networks

External signals excite a chain of response reactions in biological systems, finally leading to an adaptation of the biological system to accommodate to its changing environ-